

(12) **EUROPEAN PATENT APPLICATION**

(43) Date of publication:  
**02.05.2001 Bulletin 2001/18**

(51) Int Cl.7: **H01L 29/423, G11C 16/04**

(21) Application number: **00870245.8**

(22) Date of filing: **25.10.2000**

(84) Designated Contracting States:  
**AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU**  
**MC NL PT SE**  
 Designated Extension States:  
**AL LT LV MK RO SI**

(72) Inventors:  
 • **van Houdt, Jan**  
**3460 Bekkevoort (BE)**  
 • **Xue, Gang**  
**Wu Xi 214031 (CN)**

(30) Priority: **25.10.1999 US 161275 P**

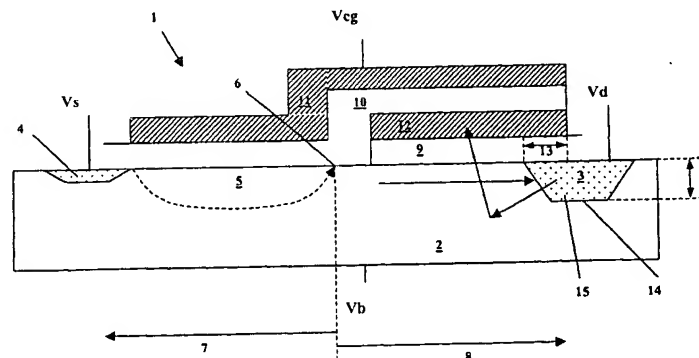
(74) Representative: **Quintelier, Claude et al**  
**Gevers & Vander Haeghen,**  
**Livornostraat 7**  
**1060 Brussels (BE)**

(71) Applicant: **INTERUNIVERSITAIRE**  
**MICROELEKTRONICA CENTRUM VZW**  
**3001 Leuven (BE)**

(54) **Electrically programmable and erasable memory device and method of operating same**

(57) An electrically programmable and erasable memory device, comprising at least one transistor (1) comprising a substrate (2) provided with a source (4), a drain (3) and a channel region (3) extending between said source and said drain. Said substrate (2) has a split point (6) situated between said source (4) and said drain (3) which forms a separation between a first region (7) extending from said split point in a first direction towards said source and a second region (8) extending from said split point in a second direction towards said drain. A first insulating layer (9) is applied on said substrate and extends in said second region (8) over at least a portion of said drain (3) and at least a portion of said channel region (5). A second insulating layer (10) is applied on

said substrate (2) in said first region (7) and separates said substrate in said first region from a control gate (11). Said second insulating layer (10) extends further in said second region (8) and contacts said control gate (11). A floating gate (12) is sandwiched between said first insulating layer (9) and said second insulating layer (10) and extends over at least a portion of said drain (3) to establish an overlap (13) between said floating gate and said drain. Said first insulating layer (9) and said overlap (13) are dimensioned in such a way as to create a capacitive coupling between said floating gate (12) and said drain (3) enabling injection onto the floating gate of hot electrons generated by drain induced secondary impact ionisation.



**Fig. 1**

## Description

### FIELD OF THE INVENTION

[0001] The invention relates to an electrically programmable and erasable memory device, comprising at least one transistor comprising:

a substrate provided with a source, a drain and a channel region extending between said source and said drain, said substrate having a split point situated between said source and said drain and forming a separation between a first region extending from said split point in a first direction towards said source and a second region extending from said split point in a second direction towards said drain; a first insulating layer applied on said substrate and extending in said second region over at least a portion of said drain and at least a portion of said channel region;  
a second insulating layer applied on said substrate in said first region and separating said substrate in said first region from a control gate, said second insulating layer extending further in said second region and contacting said control gate;  
and a floating gate sandwiched between said first insulating layer and said second insulating layer and extending over at least a portion of said drain to establish an overlap between said floating gate and said drain.

### BACKGROUND OF THE INVENTION

[0002] An EEPROM device is for example known from US-A-5,572,054. This document describes an electrically programmable and erasable memory device which comprises at least one transistor. This transistor comprises a substrate which is provided with a source, a drain and a channel region extending between the source and the drain. The substrate has a split point situated between the source and the drain which forms a separation between a first region extending from the split point towards the drain and a second region extending from the split point towards the source. A first insulating layer is applied on the substrate and extends in the second region over a portion of the source and the channel region. A second insulating layer is applied on the substrate in the first region, where it separates the substrate from a control gate. The second insulating layer further extends in the second region where it contacts the control gate. A floating gate is sandwiched between the first and second insulating layers and extends over a portion of the source to be capacitively coupled to the source. This transistor structure is commonly known in the art as a "split gate" structure.

[0003] There are n-channel and p-channel devices with split gate transistors. In the n-channel devices, the source and drain are doped with an n-type dopant and

the substrate is doped with a p-type dopant. In p-channel devices, the source and drain are doped with a p-type dopant and the substrate is doped with an n-type dopant. The device described in US-A-5,572,054 is an n-channel device. This implies that electrons flow through the channel region from the drain towards the source. In p-channel devices the electrons flow from source to drain, which implies that in p-channel split gate transistors the floating gate is located in the region extending from the split point towards the drain.

[0004] The floating gate of the device described in US-A-5,572,054 can be charged to obtain a programmed state and discharged to obtain a non-programmed or erased state of the memory cell. Programming the floating gate means that electrons are introduced onto the floating gate. Erasing means that electrons are removed from the floating gate. Assuming that the floating gate is in an erased state, i.e. positively charged, programming the transistor, i.e. charging the floating gate, is conducted as follows. A ground potential is applied to the drain, a low positive voltage (e.g. +1 V) is applied to the control gate and a high positive voltage (e.g. +12 V) is applied to the source. The high voltage difference between drain and source causes electrons to migrate through the channel from the drain towards the source, i.e. the channel region becomes conductive and is "turned on". The positive voltage on the control gate serves to transfer the drain potential onto the split point. When the electrons reach the split point, they see a steep potential drop as the influence of the positive voltage on the control gate diminishes in this point. The steep potential drop is approximately equal to the source potential and causes them to be accelerated or "heated". Due to the capacitive coupling with the source, the floating gate attracts the heated electrons, which causes some of them to be injected through the first insulating layer onto the floating gate. This process continues until the positive charges on the floating gate are neutralised by the electrons injected onto it and the floating gate is no longer positively charged, which results in the portion of the channel region beneath the floating gate being "turned off", i.e. it is no longer conductive. This method of charging the floating gate is commonly known in the art as channel hot electron injection (CHEI).

[0005] Assuming that the floating gate is in a programmed state, i.e. negatively charged, erasing the transistor, i.e. discharging the floating gate is conducted as follows. A ground potential is applied to the source and the drain, and a high positive voltage (e.g. +15 V) is applied to the control gate. The high potential of the control gate causes electrons on the floating gate to travel through the second insulating layer to the control gate by means of the Fowler-Nordheim tunneling mechanism, which is known to the person skilled in the art.

[0006] The memory cell described in US-A-5,572,054 however has the disadvantage that high voltages are needed for both programming and erasing the memory

cell. Because of these high voltages, the first and second insulating layers need to have a substantial thickness in order to avoid breakdown. Furthermore, particular circuits, such as for example charge pumping circuits, are required to achieve the high programming and erasing voltages, since these voltages are above the supply voltage of the device, which is commonly about 5 volts. This can lead to an increase in the size of the memory device.

#### AIM OF THE INVENTION

**[0007]** It is an aim of the invention to present an electrically programmable and erasable memory device in which the voltages used for programming and erasing are less than those used in the prior art.

#### DESCRIPTION OF THE INVENTION

**[0008]** The aim of the invention is achieved in that said first insulating layer and said overlap are dimensioned in such a way as to create a capacitive coupling between said floating gate and said drain enabling injection onto the floating gate of hot electrons generated by drain induced secondary impact ionisation.

**[0009]** The mechanism used in the device of the invention for programming the transistor, drain induced secondary impact ionisation, can be explained as follows. A voltage difference is applied over the channel region in such a way that hot electrons flow from source to drain. As these hot electrons impact on the drain, they transfer a certain amount of their energy onto the drain. As a result, the drain is ionised, meaning that electrons come loose from the drain. These so-called "secondary electrons" are heated as they receive energy from the electrons impacting on the drain. Due to the capacitive coupling of the floating gate with the drain, by which part of the voltage on the drain is induced on the floating gate, the secondary electrons are attracted by the floating gate. Some of them have sufficient energy to diffuse through the first insulating layer and be injected onto the floating gate. As will be explained in detail below, the mechanism of drain induced secondary impact ionisation allows the programming and erasing of the transistor at more moderate voltages with respect to the prior art.

**[0010]** It should be noted that the device of the invention is a p-channel device, which means that the drain and the source are switched with respect to the device described in US-A-572,054.

**[0011]** The mechanism of drain induced secondary impact ionisation is known as such from US-A-5,659,504. However, the transistor with which the mechanism of drain induced secondary impact ionisation is used in this document has a different structure than the transistor in the device of the invention. The transistor in the device described in US-A-5,659,504 has a so-called "stacked gate" structure. This means that the

floating gate and the control gate are stacked above each other, the floating gate being separated from the substrate by a first insulating layer and the control gate being separated from the floating gate by a second insulating layer. The floating gate and the control gate have substantially the same length and extend over the channel region between source and drain. A first main difference with the split gate structure is the absence of the split point. A second main difference is that the floating gate does substantially not extend over a portion of the drain, which means that the floating gate is substantially not capacitively coupled to the drain.

**[0012]** In the device of US-A-5,659,504, the electrons migrating through the channel region from source to drain are heated by means of the voltage difference between the drain and the source. This voltage difference has to be large enough to heat the electrons sufficiently and enable them to impact ionise on the drain. There is no indication in US-A-5,659,504 that a steep potential drop as the result of a split point can be used for sufficiently heating the electrons.

**[0013]** Furthermore, in the device of US-A-5,659,504, the floating gate is substantially not capacitively coupled to the drain, so that substantially no part of the drain voltage is induced on the floating gate. This means that substantially no injection of secondary electrons onto the floating gate can be achieved as the result of a capacitive coupling of the floating gate with the drain. The injection of secondary electrons is achieved by applying a voltage to the control gate which is such that it established an electric field attracting the secondary electrons towards the floating gate. There is no indication in US-A-5,659,504 that injection of secondary electrons onto the floating gate can be induced by capacitively coupling the floating gate to the drain. Hence, there is no indication in US-A-5,659,504 that the mechanism of drain induced secondary impact ionisation can be applied for programming a split gate transistor.

**[0014]** In the device of US-A-5,572,054 the electrons which are injected onto the floating gate are electrons which migrate through the channel region and become heated when they see the steep potential drop as a result of the split point. These electrons could be termed "primary electrons", as they are directly injected from the channel region onto the floating gate. These primary electrons are not generated on the source (or the drain) by means of impact ionisation, which means that they are not secondary electrons. There is no indication in US-A-5,572,054 that secondary electrons generated by impact ionisation of the source (or the drain) could be injected onto the floating gate of a split gate transistor. Furthermore, there is no indication in US-A-5,572,054 that primary electrons can be used to generate secondary electrons on the source (or the drain) by means of impact ionisation. Hence, there is no indication in US-A-5,572,054 that a split gate transistor can be programmed by using the mechanism of drain induced secondary impact ionisation.

[0015] It can be concluded that the device of the invention cannot be achieved by simply combining the split gate structure of US-A-5,572,054 with the programming mechanism of US-A-659,504.

[0016] In a preferred embodiment of the device of the invention, the substrate is provided to be negatively biased with respect to the source during programming of the transistor. Biasing the substrate negatively with respect to the source has the advantage that the electric field which is present over the first insulating layer and is caused by the voltage difference between the floating gate and the substrate, can be enhanced. An enhancement in this electric field causes the secondary electrons to be more attracted to the floating gate. As a result, biasing the substrate negatively with respect to the source can lead to an enhancement of the programming speed.

[0017] In a further preferred embodiment of the device of the invention, a drain junction is provided between the drain and the substrate, which drain junction has a depth larger than the overlap between the floating gate and the drain. This deep drain junction is preferably provided with a halo extension. By providing such a drain junction, the mechanism of drain induced secondary impact ionisation can be enhanced, resulting in a further enhancement of the programming speed.

[0018] The capacitive coupling between the floating gate and the drain is preferably constructed such that it enables tunnelling, preferably Fowler-Nordheim tunnelling, of electrons from said floating gate to said drain for erasing the transistor. In order to enable tunnelling of electrons from the floating gate and a target, a capacitive coupling between the floating gate and the target is required. This capacitive coupling is preferably between predetermined values. A capacitive coupling of too low value is undesirable for tunnelling, because this implies that there is either substantially no overlap between the floating gate and the target, or that the insulating layer between the floating gate and the target is too thick to enable tunnelling at a moderate voltage. A capacitive coupling of too high value is also undesirable for tunnelling, because a high capacitive coupling results in a large part of the voltage applied to the target being induced on the floating gate, so that at a moderate voltage, the voltage difference between the target and the floating gate remains too low to achieve tunnelling.

[0019] In the prior art, erasing the floating gate is achieved by Fowler-Nordheim tunnelling from the floating gate to the control gate. In the device of US-A-5,572,054, tunnelling from the floating gate to the source is not possible at a moderate voltage, because the capacitive coupling between the floating gate and the source is too high (80%). In the device of US-A-5,569,504, tunnelling from the floating gate to the drain is not possible at a moderate voltage, because there is substantially no overlap between the floating gate and the drain. In both prior art devices, the capacitive coupling between the floating gate and the control gate is

more desirable for tunnelling than the capacitive coupling between the floating gate and the source or the drain. As a result, in both prior art devices the floating gate is erased by means of tunnelling of electrons from the floating gate to the control gate. It should be noted that in both cases the voltage applied to the control gate for erasing the floating gate is still high (15 V in US-A-5,572,054; 12 to 20 V in US-A-5,569,504) with respect to the supply voltage (e.g. 5 V).

[0020] In the device of the invention, tunnelling of electrons from the floating gate to the drain is enabled, because of a suitable capacitive coupling between the floating gate and the drain (e.g. 20 to 50%). The suitable capacitive coupling results from the use of drain induced secondary impact ionisation as mechanism for programming. Because of the lower drain voltage with respect to the prior art, the first insulating layer between the floating gate and the drain can be constructed thinner. Because of the thinner first insulating layer, a smaller part of the drain voltage has to be induced on the floating gate to enable injection of secondary electrons through the first insulating layer. This means that the capacitive coupling between the floating gate and the drain can have a lower value than in the device of US-A-5,572,054. In the latter device, the floating gate has to be induced to a higher voltage value, due to use of channel hot (primary) electron injection as mechanism for programming the floating gate. This is because a high source voltage is required to sufficiently heat the primary electrons, which in turn results in the requirement of a thicker first insulating layer between the floating gate and the source in order to prevent breakdown, which in turn leads to the primary electrons needing a higher amount of energy to cross the first insulating layer, so that the floating gate is to be induced to a higher voltage value. The presence of a suitable capacitive coupling between the floating gate and the drain in the device of the invention also leads to a lower voltage (e.g. about 8 V) being required on the drain for erasing the floating gate, with respect to the voltages needed on the control gate in the prior art.

[0021] Programming the transistor of the device of the invention comprises the steps of applying a first source voltage to the source, applying a first control gate voltage to the control gate and applying a first drain voltage to the drain. The first drain voltage has a higher voltage value than the first control gate voltage, which in its turn has a higher voltage value than the first source voltage. As the first control gate voltage is below the first drain voltage, the device of the invention allows the use of more moderate voltages for programming with respect to the prior art.

[0022] Erasing the transistor of the device of the invention comprises the steps of applying a second source voltage to the source, applying a second control gate voltage to the control gate and applying a second drain voltage to the drain. The second drain voltage has a higher voltage value than the second control gate volt-

age and the second source voltage, which are preferably supplied with the ground potential. As a result of the suitable capacitive coupling between the floating gate and the drain as described above, the device of the invention allows the use of more moderate voltages for erasing with respect to the prior art.

[0023] Reading the transistor of the device of the invention comprises the steps of applying a third source voltage to the source, applying a third control gate voltage to the control gate and applying a third drain voltage to the drain. The third control gate voltage has a higher voltage value than the third source voltage, which in its turn has a higher voltage value than the third drain voltage. This method of reading the transistor can be termed "reverse read-out", because the third source voltage is higher than the third drain voltage, which is preferably the ground potential. The reverse read-out has the advantage that a low voltage, preferably the ground potential, is applied to the drain during reading, which serves to avoid a leakage current from the floating gate to the drain. The reverse read-out is in contrast to the prior art, because for example in the device of US-A-5,572,054, the higher voltage for reading is applied on the floating gate side of the transistor, i.e. also on the source, but as already mentioned, the drain and source are switched in the prior art device with respect to the device of the invention.

[0024] Further advantages of the device according to the invention will appear from the following description and the appended figures.

#### BRIEF DESCRIPTION OF THE DRAWINGS

[0025] Figure 1 shows a cross-sectional side view of a preferred embodiment of the transistor of the device of the invention.

[0026] Figure 2 shows a graph representing the influence of negatively biasing the substrate on the floating gate current in the device of the invention.

[0027] Figure 3 shows a suitable array configuration for the device of the invention.

#### DETAILED DESCRIPTION

[0028] The device shown in figure 1 comprises a transistor 1 which comprises a substrate 2 provided with a drain 3, a source 4 and a channel region 5 extending between the source 4 and the drain 3. The substrate 2 has a split point 6 situated between the source 4 and the drain 3 and forming a separation between a first region 7 extending from the split point 6 in a first direction towards the source 4 and a second region 8 extending from the split point 6 in a second direction towards the drain 3. A first insulating layer 9 is applied on the substrate 2 and extends in the second region 8 over at least a portion of the drain 3 and at least a portion of the channel region 5. A second insulating layer 10 is applied on the substrate 2 in the first region 7, where it separates

the substrate 2 from a control gate 11. The second insulating layer 10 extends further in the second region 8, where it contacts the control gate 11. A floating gate 12 is sandwiched between the first insulating layer 9 and the second insulating layer 10. The floating gate 12 extends in the second region 8 over the channel region 5 and over at least a portion of the drain 3 to establish an overlap 13 between the floating gate 12 and the drain 3. The first insulating layer 9 and the overlap 13 are dimensioned in such a way as to create a capacitive coupling between the floating gate 12 and the drain 3. This capacitive coupling enables injection onto the floating gate 3 of hot electrons generated by drain induced secondary impact ionisation.

[0029] In the device of the invention, the first insulating layer 9 and the overlap 13 between the floating gate 12 and the drain 3 are dimensioned in such a way as to create a capacitive coupling between the floating gate 12 and the drain 3 enabling injection onto the floating gate 3 of hot electrons generated by drain induced secondary impact ionisation. Primary electrons, i.e. hot electrons migrating through the channel region 5 from source 4 to drain 3, require less energy for impact ionising the drain 3 than they require for being injected onto the floating gate 12. This is because the injection onto the floating gate 12 requires a lot of energy to be able to cross the first insulating layer 9. As a result, a lower voltage difference between drain 3 and source 4 is required for heating the primary electrons when they are used for impact ionisation of the drain 3 instead of injection onto the floating gate 12. This implies that, in the device of the invention, only a moderate voltage needs to be supplied to the drain 3 for heating the primary electrons. In turn, this allows a thinner first insulating layer 9, as the voltage difference between the drain 3 and the floating gate 12 will never be as large as when the primary electrons are used for injection onto the floating gate 12. The first insulating layer 9 can be constructed thinner as there is less risk of breakdown. Because the first insulating layer 9 can be constructed thinner, the amount of energy required for injection of hot electrons onto the floating gate 12 is reduced. This in turn allows the use of secondary electrons for charging the floating gate 12 in the device of the invention. So the device of the invention has the advantage that the voltage applied to the drain 3 for programming the transistor can be reduced with respect to the voltage applied to the source of the device of US-A-5,572,054.

[0030] By providing the split point 6, the electrons migrating through the channel region 5 can become sufficiently heated by the steep potential drop to enable them to impact ionise on the drain 3. This implies that a lower voltage with respect to the source 4 can be applied to the drain 3. This has the advantage that the voltage difference over the channel region 2 can be reduced with respect to the device of US-A-5,659,504.

[0031] By providing the capacitive coupling between the floating gate 12 and the drain 3, part of the drain

voltage is induced on the floating gate 12, enabling the floating gate 12 to attract the secondary electrons. As a result, the control gate voltage is no longer used for attracting the secondary electrons towards the floating gate 12; it is only used for biasing the channel region 5 in said first region 7 extending from the split point 6 towards the source 4, in such a way that the source voltage is transferred onto the split point 6. As the control gate 11 in the first region 7 is only separated from the substrate 2 by the second insulating layer 10, and not also by the first insulating layer 9 and the floating gate 12 as in the stacked gate transistor, i.e. the control gate 11 is nearer the substrate 2 in the first region 7, providing the capacitive coupling has the advantage that the control gate voltage required for programming the transistor 1 can be reduced with respect to the device of US-A-5,659,504. Furthermore, the voltage required on the control gate 11 for programming the transistor in the device of the invention can be below the voltage applied to the drain 3.

[0032] The device of the invention can be operated at lower voltages than prior-art devices and thus consumes less power, is smaller in size, is more scalable, requires less charge pumping circuitry. Furthermore, in the device of the invention, the use of a program gate for triggering the floating gate can be avoided, as the floating gate is capacitively coupled to the drain. Such a program gate is for example required in prior-art devices using Source Side Injection at moderate voltages. The omission of the program gate enables the construction of a smaller memory device as compared to such Source Side Injection based devices. The transistor size in the device of the invention can be less than  $1 \mu\text{m}^2$  in a  $0.25 \mu\text{m}$  CMOS technology.

[0033] The capacitive coupling ratio of the floating gate 12 with respect to the drain 3 is preferably between 0.2 and 0.5. This means that preferably 20 to 50% of the voltage applied to the drain is induced on the floating gate. The coupling ratio between the floating gate and the drain can however also be any other value deemed suitable by the person skilled in the art.

[0034] The substrate 2 is preferably provided to be negatively biased with respect to the source 4 during programming of the transistor 1. The effect of negatively biasing the substrate 2 is that the electric field which is created over the first insulating layer 9, i.e. between the floating gate 12 and the substrate 2 is enhanced. This results in the secondary electrons being more strongly attracted by the floating gate 12, so that more secondary electrons are injected onto the floating gate 12 in a given period. So by negatively biasing the substrate 2 with respect to the source 4, the programming speed of the device according to the invention can be enhanced.

[0035] The enhancement in the programming speed is illustrated in figure 2, which represents the floating gate current  $I_{fg}$  for charging the floating gate as a function of the floating gate voltage  $V_{fg}$  (with the source being connected to the ground potential) for a zero sub-

strate bias  $V_b$  and a negative substrate bias  $V_b$  of  $-2.5 \text{ V}$ , and for a device produced in  $0.25 \mu\text{m}$  CMOS technology. When grounding the substrate 2, only a very small floating gate current  $I_{fg}$  is detected because of the poor injection efficiency of the conventional drain hot-electron injection mechanism. However, when a small negative voltage is applied to the substrate (e.g.  $-2.5 \text{ V}$ ), the gate current  $I_{fg}$  is increased by several orders of magnitude due to secondary electron injection effects originating from a larger silicon electric field in the drain region. This experiment evidences the appearance of an injection mechanism in the memory device according to the invention, that can be used for fast programming at low voltages.

[0036] As the memory device of the invention preferably comprises a plurality of transistors, arranged in parallel columns and rows, the substrate 2 is preferably locally adapted for ensuring electrical isolation of each transistor for which the substrate is negatively biased with respect to the source 4, from the rest of the substrate. In this way it can be ensured that transistors which do not have to be programmed, i.e. for which the substrate is not to be negatively biased with respect to the source, are unintentionally programmed.

[0037] The device of figure 1 is preferably provided with a drain junction 14 having a depth  $D$  which is optimised for having a highly efficient drain induced secondary impact ionisation. The optimised depth can be achieved by making the drain junction depth  $D$  is larger than the overlap 13 between the floating gate 12 and the drain 3. The drain junction depth  $D$  is preferably between one to four times the overlap 13, or larger. The drain junction 14 is further preferably provided with a halo extension in order to further increase secondary electron injection efficiency. A large drain junction depth  $D$  is in contrast with the prior art devices, in which a shallow drain is provided in order to increase the electric field between the drain and the source. The large drain junction depth  $D$  is possible in the device according to the invention, as the electric field between the drain and the source need not be as strong as in the prior art devices.

[0038] The device according to the invention shows a programming efficiency which is at least similar to prior-art devices but at much lower voltages due to the use of the drain induced secondary impact ionisation mechanism for programming, which requires a drain voltage which can be less than the supply voltage to the device of for example 5 volts. The low drain voltage allows the use of a much thinner first insulating layer 9 under the floating gate 12 than in prior art devices, since drain disturb conditions are largely relaxed. This in turn enables erase of the floating gate 12 towards the drain 3 by means of tunnelling of electrons through the first insulating layer 9, instead of erasing the floating gate towards the control gate by means of tunnelling of electrons through the second insulating layer 10. As a result, only low voltages are to be applied to the control gate 11, both during programming and erasing of the transis-

tor 1. Consequently, also the second insulating layer 10 under the control gate 11 can be scaled in relation to the corresponding CMOS generation, i.e. can be constructed thinner with respect to the prior art, as there is less risk of breakdown of the second insulating layer 10 resulting from a high voltage on the control gate 11. A second reason why the second insulating layer 10 is to remain very thick in prior art devices, for example split gate devices, is the need for a very large drain coupling to enable injection of primary hot electrons onto the floating gate. As the sum of the respective coupling ratios between the floating gate and the respective components of the transistor surrounding the floating gate equals 1 (by definition), this implies that the coupling ratio between the floating gate 12 and the control gate 11 should be minimised. In the present invention, the control gate coupling is allowed to be larger because a drain coupling on the order of 20%-50% is sufficient to induce enough voltage on the floating gate to enable the injection of secondary electrons onto the floating gate, which allows a thinner second insulating layer 10.

[0039] The first insulating layer 9 preferably has a thickness of at most 50 angstroms (5 nm). The second insulating layer 10 in the first region 7 preferably has a thickness of at most 50 angstroms (5 nm), preferably 35 angstroms (3.5 nm). The second insulating layer 10 in the second region 8 preferably has a thickness of at most 150 angstroms (15 nm), preferably 130 angstroms (13 nm). The first and second insulating layers can however also have any thickness deemed suitable by the person skilled in the art. The first and second insulating layer can have the same or different dielectric constants.

[0040] An exemplary processing scheme for such a device comprises the following steps: (1) after growing a thin oxide on the substrate 2 in the second region 8 to form the first insulating layers 9 (typically 70Å for a 0.35µm CMOS technology), a first polysilicon layer is deposited and etched to form the floating gates 12 of the transistors 1. Secondly, a junction 14 is formed by a deep n<sup>+</sup> implantation (preferably combined Phosphorous/Arsenicum junction with a halo) which is self-aligned with respect to the floating gate 12 on the drain side of the device. The source junction 4 could be formed by the same implantation step, i.e. in a non-self-aligned manner, or, alternatively, it can also be formed together with the CMOS junctions (after 2nd polysilicon definition). Afterwards, a thin oxide (comparable to the CMOS gate oxide of the corresponding generation, i.e. 55Å for 0.25µm CMOS etc.) is grown on the complementary part of the substrate 2, i.e. in the first region 7, and simultaneously a thin polyoxide is formed on the top and the sidewall of the floating gate 12, to form the second insulating layer 10. Depending on the oxidation conditions and the doping level of the floating gate 12, this interpoly oxide can be very thin. Then, a second polysilicon layer is deposited and etched to form the control gate 11 of the transistor 1. At this point, the junctions of the CMOS process are formed which can, eventually,

be combined with the source junctions of the transistors. The device of the invention may however also be produced in any other way known to the person skilled in the art.

[0041] Programming the device of figure 1 is carried out according to the following method. A first source voltage  $V_{s1}$  is applied to the source 4, a first control gate voltage  $V_{cg1}$  to the control gate 11 and a first drain voltage  $V_{d1}$  to the drain 3. The first drain voltage  $V_{d1}$  has a higher voltage value than the first control gate voltage  $V_{cg1}$ , which in its turn has a higher voltage value than the first source voltage  $V_{s1}$ . The first source voltage  $V_{s1}$  is preferably the ground potential. The first drain voltage  $V_{d1}$  is preferably below the supply voltage to the device of for example 5 V. The voltage difference between the first drain voltage  $V_{d1}$  and the first source voltage  $V_{s1}$  is above the threshold voltage  $V_t$  for turning on the channel region 5. The first control gate voltage  $V_{cg1}$  is in such a way above the first source voltage  $V_{s1}$  that the first source voltage  $V_{s1}$  is transferred onto the split point 6. The first drain voltage  $V_{d1}$  is further chosen such that a high enough voltage is induced on the floating gate 12, as a result of the capacitive coupling with the drain 3, that injection of secondary electrons, which are generated by impact ionisation on the drain, is enabled.

[0042] Preferably also a substrate voltage  $V_b$  is applied to the substrate 2, which is preferably negative with respect to the first source voltage  $V_{s1}$ . This substrate voltage  $V_b$  serves to increase the electric field over the first insulating layer 9, so that the injection of secondary electrons onto the floating gate 12 can be enhanced.

[0043] Typical programming voltages for the device of the invention in a 0.18 µm technology are: a first source voltage  $V_{s1}$  of 0 V (the source is grounded), a first control gate voltage  $V_{cg1}$  of around 2 V and a first drain voltage  $V_{d1}$  of 4-5V which can be supplied from a small charge pumping circuit. A small negative substrate voltage  $V_b$  of about -2V or less is preferably applied to the substrate. This brings the floating gate 12 to a potential of about 3V which is sufficient to efficiently trigger the drain enhanced secondary impact ionisation mechanism.

[0044] Erasing the device of figure 1 is carried out according to the following method. A second source voltage  $V_{s2}$  is applied to the source 4, second control gate voltage  $V_{cg2}$  to the control gate 11 and a second drain voltage  $V_{d2}$  to the drain 3. The second drain voltage  $V_{d2}$  has a higher voltage value than the second source voltage  $V_{s2}$  and the second control gate voltage  $V_{cg2}$ . The second drain voltage  $V_{d2}$  is preferably above the supply voltage and the are preferably below the supply voltage applied to the device. The second source and control gate voltages  $V_{s2}$  and  $V_{cg2}$  are preferably the ground potential. The second drain voltage  $V_{d2}$  is chosen in such a way above the second control gate voltage  $V_{cg2}$  that electrons on the floating gate 12 are transferred to the drain 3 by means of tunnelling, preferably Fowler-Nordheim tunnelling, through the first insulating layer 9.

[0045] During erase, the control gate 11 is preferably



grounded while the second drain voltage  $V_{d2}$  is preferably about 8V. A small negative voltage could be applied to the substrate 2 to further increase the tunneling field at the floating-gate-to-drain overlap 13.

[0046] Reading the device of figure 1 is carried out according to the following method. A third drain voltage  $V_{d3}$  is applied to the drain 3, a third source voltage  $V_{s3}$  to the source 4 and a third control gate voltage  $V_{cg3}$  to the control gate 11. The third control gate voltage  $V_{cg3}$  has a higher voltage value than the third source voltage  $V_{s3}$ , which in turn has a higher voltage value than the third drain voltage  $V_{d3}$ . The third control gate voltage  $V_{cg3}$  is preferably below the supply voltage to the device and the third drain voltage  $V_{d3}$  is preferably the ground potential. As already mentioned above, this method of reading the transistor can be termed "reverse read-out", as the voltage for reading is applied to the source 4 instead of to the drain 3. This has the advantage that the leakage current from floating gate 12 to drain 3 is suppressed. This enhances the reliability of the device of the invention.

[0047] Due to the possibility of using thin insulating layers 9, 10 under control gate 11 and floating gate 12, the device has also a high read-out current. This also implies that the programmed state is a 'hard-off' state since the subthreshold slope of the transistor has a steepness comparable to the CMOS devices in the same technology. For read-out a source voltage  $V_{s3}$  being 1 V higher than the (grounded) drain voltage  $V_{d3}$  can be used. A read-out control gate voltage  $V_{cg3}$  in between 1.8 and 2.5 higher than the (grounded) drain voltage  $V_{d3}$  can be exploited. The substrate 2 is preferably also grounded.

[0048] Figure 3 shows an efficient array organisation for the memory device of the invention cell when fabricating memory circuits. First, it is noted that the sources 4 of the transistors on a column are to be connected to the vertical bitline, while the drains 3 of the cells on a row are connected to a common horizontal erase line. Advantages of this configuration are that the cell is read-out in the reverse way, which suppresses the Stress-Induced Leakage Current in the drain-to-floating gate overlap region. Additionally, the absence of drain coupling during read-out further reduces the amount of electrons to be transferred onto the floating gate 12 for a given external threshold voltage  $V_t$  and, thus also the electric field over the first insulating layer 9 under charge storage (or retention) conditions. Secondly the high erase voltage is only applied to one particular row of cells (or, eventually to a number of adjacent rows) which are to be erased simultaneously as a sector. This ensures that the erase voltage  $V_{d2}$  does not disturb the other sectors of the memory (no erase disturb mechanism and thus no need for inhibit voltages). Thirdly the moderate drain voltage  $V_{d1}$  applied during programming will not cause significant charge loss in unselected transistors since the disturb time is limited by the number of words on a row. Since the drain 3 is connected to a com-

mon erase line along a row of cells, the non-selected cells have to be inhibited during programming, i.e. prevented from being programmed unintentionally. This can easily be done as follows: (1) all bitlines are biased at the supply voltage or a slightly larger inhibit voltage (e.g. 2.5 V in a 0.1  $\mu\text{m}$  technology); (2) a row of cells is selected by applying about 1.8V to its wordline and 4-5V to its erase line. Note that none of the transistors is drawing any current under these conditions since the control-gate channel, i.e. the channel region 2 in the first region 7 under the control gate 11, is cut off. (3) the bitlines of the transistors to be programmed are discharged selectively to ground which causes a current to flow only through these cells. A consequence of this configuration is that the bitline cannot be shared between adjacent columns of transistors. It is, however, possible to share the bitline contact between 2 adjacent transistors on the same column in order to reduce the transistor area. For the drain contact (contacting the transistor to the erase line), the situation is somewhat more complicated. If the drain contact is shared between adjacent transistors on the same column, the impact of the drain disturb mechanism during programming is more than doubled because of the absence of a wordline voltage on the adjacent row, which further enhances the tunneling field across the first insulating layer 9. In practice, this problem requires a compromise between sector size, first insulating layer thickness and drain voltage during programming. Sharing the drain contact between transistors on the same row, or, alternatively, using a diffusion region for erase line routing are other solutions that remove the disturb problem.

[0049] The possible voltages for programming, erasing and reading a transistor in the device of the invention are summarised in the table below.

	$V_s(\text{V})$	$V_d(\text{V})$	$V_{cg}(\text{V})$	$V_b(\text{V})$
program	0	4-5	1.8-4	~-2
read-out	1	0	1.8-2.5	0
erase	0	8	0	0

#### Claims

1. An electrically programmable and erasable memory device, comprising at least one transistor (1) comprising:

a substrate (2) provided with a source (4), a drain (3) and a channel region (3) extending between said source and said drain, said substrate (2) having a split point (6) situated between said source (4) and said drain (3) and forming a separation between a first region (7) extending from said split point in a first direction towards said source and a second region (8)



extending from said split point in a second direction towards said drain;  
 a first insulating layer (9) applied on said substrate and extending in said second region (8) over at least a portion of said drain (3) and at least a portion of said channel region (5);  
 a second insulating layer (10) applied on said substrate (2) in said first region (7) and separating said substrate in said first region from a control gate (11), said second insulating layer (10) extending further in said second region (8) and contacting said control gate (11);  
 a floating gate (12) sandwiched between said first insulating layer (9) and said second insulating layer (10) and extending over at least a portion of said drain (3) to establish an overlap (13) between said floating gate and said drain;

characterised in that said first insulating layer (9) and said overlap (13) are dimensioned in such a way as to create a capacitive coupling between said floating gate (12) and said drain (3) enabling injection onto the floating gate of hot electrons generated by drain induced secondary impact ionisation.

2. A device according to claim 1, characterised in that that said substrate (2) is provided to be negatively biased with respect to the source (4) during programming of the transistor (1).
3. A device according to claim 1 or 2, characterised in that a drain junction (14) is provided between the drain (3) and the substrate (2), said drain junction (14) having a halo extension (15).
4. A device according to claim 3, characterised in that said drain junction (14) has a depth (D) larger than said overlap (13) between said floating gate and said drain.
5. A device according to any one of claims 1-4, characterised in that said capacitive coupling between said floating gate (12) and said drain (3) enables tunnelling of electrons from said floating gate to said drain for erasing the transistor (1).
6. A device according to claim 5, characterised in that said capacitive coupling between said floating gate (12) and said drain (3) enables Fowler-Nordheim tunnelling of electrons from said floating gate to said drain for erasing the transistor (1).
7. A device according to any one of claims 1-6, characterised in that the floating gate (12) and the drain (3) are capacitively coupled with a coupling ratio of 0.2 to 0.5.
8. A device according to any one of claims 1-7, characterised in that the second insulating layer (10) in the first region (7) has a thickness of at most 50 angstroms (5 nm), preferably 35 angstroms (3.5 nm).
9. A device according to any one of claims 1-8, characterised in that the second insulating layer (10) in the second region (8) has a thickness of at most 150 angstroms (15nm), preferably 130 angstroms (13 nm).
10. A device according to any one of claims 1-9, characterised in that the first insulating layer (9) has a thickness of at most 50 angstroms (5 nm).
11. A method of programming an electrically programmable and erasable memory device as claimed in any one of claims 1-10, said method comprising the steps of applying a first source voltage ( $V_{s1}$ ) to the source (4), applying a first control gate voltage ( $V_{cg1}$ ) to the control gate (11) and applying a first drain voltage ( $V_{d1}$ ) to the drain (3), characterised in that the first drain voltage ( $V_{d1}$ ) has a higher voltage value than the first control gate voltage ( $V_{cg1}$ ), which in its turn has a higher voltage value than the first source voltage ( $V_{s1}$ ).
12. A method according to claim 11, characterised in that the first drain voltage ( $V_{d1}$ ) is below a supply voltage which is supplied to said memory device.
13. A method according to claim 11 or 12, characterised in that the first drain voltage ( $V_{d1}$ ) is between 4 and 5 volts, the first control gate voltage ( $V_{cg1}$ ) is between 1.8 and 4 volts and the first source voltage ( $V_{s1}$ ) is a ground potential of 0 volts.
14. A method according to any one of claims 11-13, characterised in that the method further comprises the step of creating a voltage difference between said source (4) and said substrate (2).
15. A method according to claim 14, characterised in that said voltage difference is such that the substrate (2) is negatively biased with respect to the source (4).
16. A method according to claim 14 or 15, characterised in that a substrate voltage ( $V_b$ ) of about -2 volts is applied to the substrate (2).
17. A method of erasing an electrically programmable and erasable memory device as claimed in claim 5 or 6, said method comprising the steps of applying a second source voltage ( $V_{s2}$ ) to the source (4), applying a second control gate voltage ( $V_{cg2}$ ) to the control gate (11) and applying a second drain voltage ( $V_{d2}$ ) to the drain (3), characterised in that the

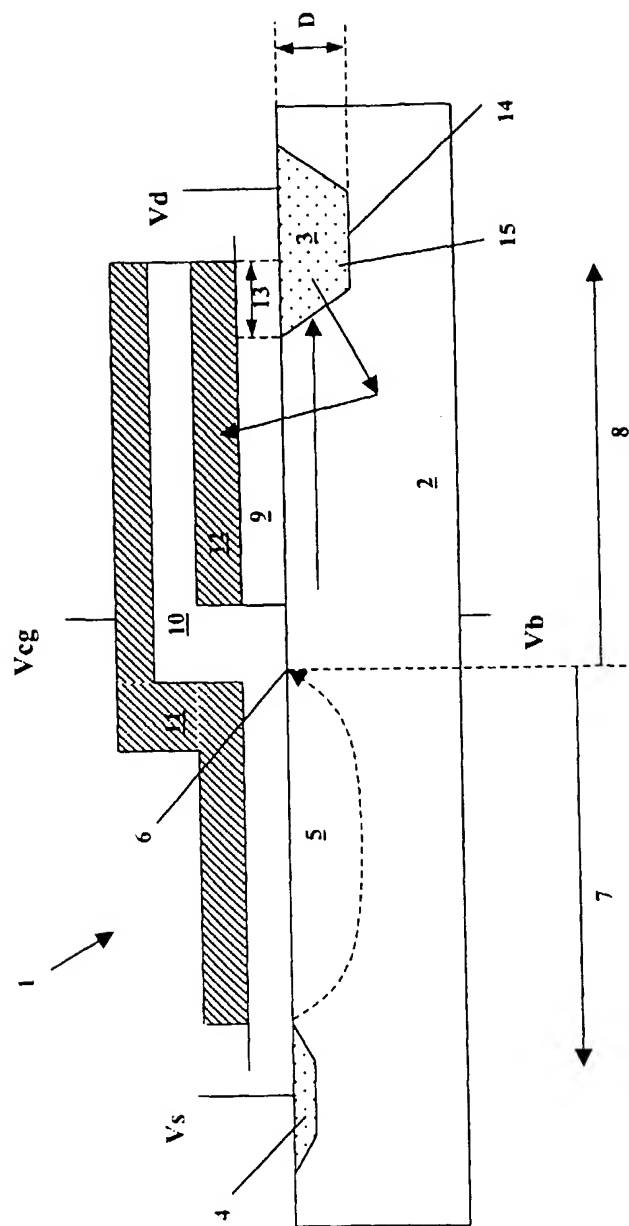
second drain voltage ( $V_{d2}$ ) has a higher voltage value than the second control gate voltage ( $V_{cg2}$ ) and the second source voltage ( $V_{s2}$ ).

18. A method according to claim 17, characterised in that a supply voltage is applied to said memory device and that the second drain voltage ( $V_{d2}$ ) is above said supply voltage and the second control gate and source ( $V_{cg2}$ ,  $V_{s2}$ ) voltages are below said supply voltage. 5 10
19. A method according to claim 17 or 18, characterised in that the second drain voltage ( $V_{d2}$ ) is about 8 volts and the second source and control gate voltages ( $V_{cg2}$ ,  $V_{s2}$ ) are the ground potential of 0 volts. 15
20. A method of reading an electrically programmable and erasable memory device as claimed in any one of claims 1-8, said method comprising the steps of applying a third source voltage ( $V_{s3}$ ) to the source (4), applying a third control gate voltage ( $V_{cg3}$ ) to the control gate (11) and applying a third drain voltage ( $V_{d3}$ ) to the drain (3), characterised in that the third control gate voltage ( $V_{cg3}$ ) has a higher voltage value than the third source voltage ( $V_{s3}$ ), which in its turn has a higher voltage value than the third drain voltage ( $V_{d3}$ ). 20 25
21. A method according to claim 20, characterised in that a supply voltage above the ground potential is applied to said memory device and that the third control gate voltage ( $V_{cg3}$ ) is between said supply voltage and the ground potential, the third source voltage ( $V_{s3}$ ) is between the third control gate voltage ( $V_{cg3}$ ) and the ground potential and the third drain voltage ( $V_{d3}$ ) is the ground potential. 30 35
22. A method according to claim 20 or 21, characterised in that the third source voltage ( $V_{s3}$ ) is about 1 volt, the third control gate voltage ( $V_{cg3}$ ) is between 1.8 and 2.5 volts and the third drain voltage ( $V_{d3}$ ) is about 0 volts. 40

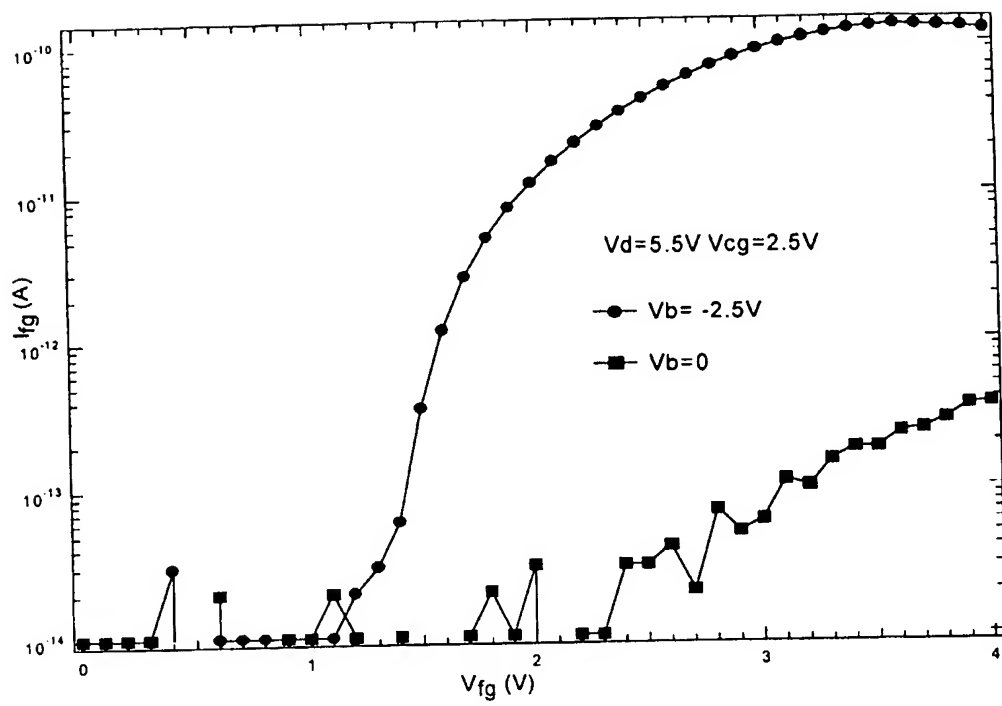
45

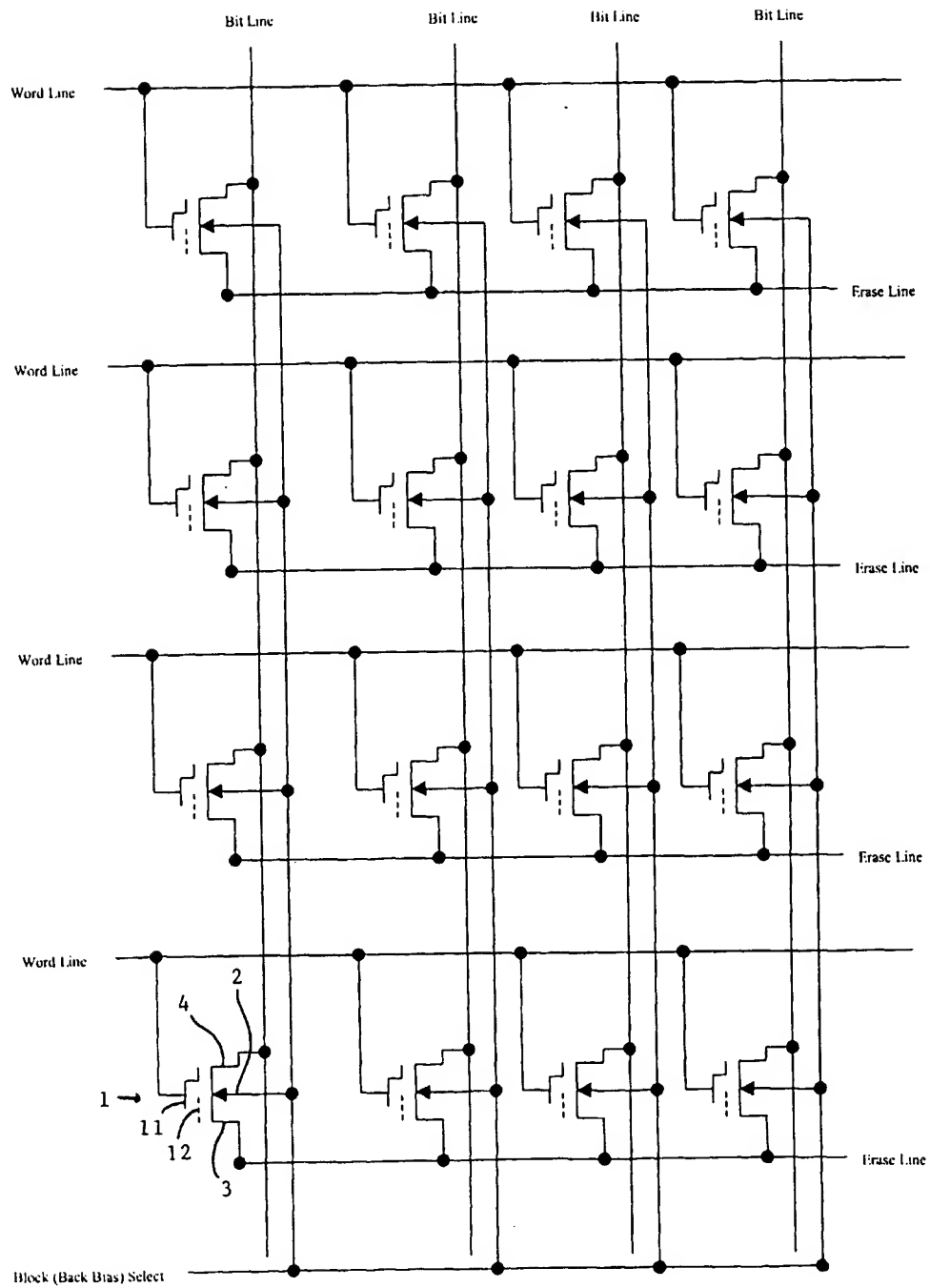
50

55



**Fig. 1**

**Fig. 2**



**Fig.3**



European Patent  
Office

## EUROPEAN SEARCH REPORT

Application Number  
EP 00 87 0245

DOCUMENTS CONSIDERED TO BE RELEVANT			
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.7)
X	EP 0 935 293 A (MATSUSHITA ELECTRIC IND CO LTD ;HALO LSI DESIGN AND DEVICE TEC (US) 11 August 1999 (1999-08-11) * column 9, line 47 - column 12, line 49; figure 1A * * column 13, line 25 - column 15, line 51 *	1-22	H01L29/423 G11C16/04
A	US 4 412 311 A (MICCOLI FRANCO ET AL) 25 October 1983 (1983-10-25) * column 3, line 41 - column 4, line 60; figures 2B,3 *	1-22	
A	US 4 361 847 A (HARARI ELIYAHOU) 30 November 1982 (1982-11-30) * column 4, line 43 - column 5, line 25; figure 2A *	1-22	
A	US 5 394 360 A (FUKUMOTO TAKAHIRO) 28 February 1995 (1995-02-28) * column 3, line 55 - column 4, line 28; figure 4 * * column 2, line 53-63 *	1-22	
The present search report has been drawn up for all claims			TECHNICAL FIELDS SEARCHED (Int.Cl.7)  H01L G11C
Place of search <b>MUNICH</b>		Date of completion of the search <b>22 February 2001</b>	Examiner <b>Lindquist, J</b>
<p>CATEGORY OF CITED DOCUMENTS</p> <p>X : particularly relevant if taken alone Y : particularly relevant if combined with another document of the same category A : technological background O : non-written disclosure P : intermediate document</p> <p>T : theory or principle underlying the invention E : earlier patent document, but published on, or after the filing date D : document cited in the application L : document cited for other reasons &amp; : member of the same patent family, corresponding document</p>			

EPO FORM 1503 03/82 (P04C01)

**ANNEX TO THE EUROPEAN SEARCH REPORT  
ON EUROPEAN PATENT APPLICATION NO.**

EP 00 87 0245

This annex lists the patent family members relating to the patent documents cited in the above-mentioned European search report. The members are as contained in the European Patent Office EDP file on  
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

22-02-2001

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 0935293 A	11-08-1999	US 6051860 A	18-04-2000
		JP 2982901 B	29-11-1999
		JP 11260944 A	24-09-1999
US 4412311 A	25-10-1983	IT 1209227 B	16-07-1989
		DE 3121753 A	03-06-1982
		FR 2484124 A	11-12-1981
		GB 2077492 A,B	16-12-1981
		JP 57052171 A	27-03-1982
US 4361847 A	30-11-1982	US 4328565 A	04-05-1982
		DE 3103160 A	21-01-1982
		GB 2073484 A,B	14-10-1981
		GB 2073487 A,B	14-10-1981
		JP 1279100 C	29-08-1985
		JP 56155574 A	01-12-1981
		JP 59051753 B	15-12-1984
		US 4409723 A	18-10-1983
US 5394360 A	28-02-1995	JP 2602575 B	23-04-1997
		JP 4065879 A	02-03-1992

EPO FORM P/459

For more details about this annex : see Official Journal of the European Patent Office, No. 12/82